

Data Mining of Patients on Weaning Trials from Mechanical Ventilation Using Cluster Analysis and Neural Networks

Carlos Arizmendi, Enrique Romero, René Alquezar, Pere Caminal, Ivan Díaz, Salvador Benito,
Beatriz F. Giraldo, *Member, IEEE*

Abstract — The process of weaning from mechanical ventilation is one of the challenges in intensive care. 149 patients under extubation process (T-tube test) were studied: 88 patients with successful trials (group S), 38 patients who failed to maintain spontaneous breathing and were reconnected (group F), and 23 patients with successful test but that had to be reintubated before 48 hours (group R). Each patient was characterized using 8 time series and 6 statistics extracted from respiratory and cardiac signals. A moving window statistical analysis was applied obtaining for each patient a sequence of patterns of 48 features. Applying a cluster analysis two groups with the majority dataset were obtained. Neural networks were applied to discriminate between patients from groups S, F and R. The best performance obtained was 84.0% of well classified patients using a linear perceptron trained with a feature selection procedure (that selected 19 of the 48 features) and taking as input the main cluster centroid. However, the classification baseline 69.8% could not be improved when using the original set of patterns instead of the centroids to classify the patients.

I. INTRODUCTION

DISCONTINUATION of mechanical ventilation, also called weaning or extubation, should be performed as soon as autonomous respiration can be sustained. It is one of the most challenging problems in intensive care units. Despite advances in mechanical ventilation and respiratory support, the science of determining if the patient

is ready for extubation is still very imprecise. A failed weaning trial is discomforting for the patient and may induce significant cardiopulmonary distress. When mechanical ventilation is discontinued, up to 25 percent of patients have respiratory distress severe enough to necessitate reinstitution of ventilatory support. Hence the need for a more accurate prediction of the optimal disconnection time, which is extended to the whole weaning process [1-2].

The variability of breathing pattern is not random and can be explained by central neural mechanisms or instability of the feedback loops [3]. The variability of the breathing pattern was previously analyzed in [4-7].

The aim of this study is to analyse respiratory pattern variability in a specific process, the weaning process, by applying both cluster analysis and neural networks, in order to find possible differences between patients who can maintain spontaneous breathing and patients who can not. It is also of interest to detect, using these methods, which variables are the most relevant for the classification.

Data mining is the process of extracting hidden patterns from data. A cluster analysis permits the classification of the similar characteristics into different groups according with the proximity distance. It is a multivariate statistical technique for unsupervised learning, especially suitable for extracting information from a data set without imposing prior restraints explicitly [8].

Neural networks are sophisticated statistical techniques capable of modelling extremely complex functions. In our case, linear and non-linear feed-forward neural networks have been used, together with feature selection methods, to classify patients who presented success or failure in the weaning process [9, 10].

II. ANALYZED DATA

Electrocardiographic signal (ECG) and respiratory flow were measured in 149 patients under mechanical ventilation and extubation process (database WEANDB). The signals were recorded from patients in Intensive Care Units of the *Hospital de la Santa Creu i Sant Pau* and the *Hospital Universitario de Getafe*, according to the protocols approved by the local ethics committees.

Manuscript received April 21, 2009. This work was supported in part by Ministerio de Ciencia e Innovación under grants TEC2007-63637, TEC2007-68076-C02-01 and TIN 2006-08114 from the Spanish Government.

C. Arizmendi, E. Romero and R. Alquezar are with Dep. of LSI, Technical University of Catalonia (UPC), C. Jordi Girona, 1-3, 08034, Barcelona, Spain (e-mail: Carlos.Julio.Arizmendi@upc.edu, eromero@lsi.upc.edu, alquezar@lsi.upc.edu).

P. Caminal and B.F. Giraldo, are with Dep. of ESAT, *Universitat Politècnica de Catalunya* (UPC), *Institut de Bioenginyeria de Catalunya* (IBEC) and *CIBER de Bioingeniería, Biomateriales y Nanomedicina* (CIBER-BBN). c/. Pau Gargallo, 5, 08028, Barcelona, Spain (e-mail: Pere.Caminal@upc.edu, Beatriz.Giraldo@upc.edu).

S. Benito and I. Díaz are with Dep. Intensive Care Medicine, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain.

Patients were classified in three groups, taking into account clinical criteria based on T-tube test: group S, 88 patients whose T-tube test was overcome successfully; group F, 38 patients who failed the test and therefore could not be extubated; and Group R, 23 patients who passed the T-tube test and were disconnected from mechanical ventilation, but they had to be reintubated before 48 hours.

The respiratory flow was obtained with a pneumotachograph (Datex-Ohmeda monitor with variable reluctance transducer) connected to an endotracheal tube. The ECG was obtained using a Spacelabs Medical monitor. Both signals were recorded during 30 minutes at a sampling frequency of 250 Hz.

The cardiac and respiratory signals were processed to obtain the following time series for each patient: cardiac interbeat duration (RR), inspiratory time (T_I), expiratory time (T_E), breath duration (T_{Tot}), tidal volume (V_T), inspiratory fraction (T_I / T_{Tot}), mean inspiratory flow (V_T / T_I) and frequency-tidal volume ratio (f / V_T).

III. METHODOLOGY AND RESULTS

A. Data preprocessing and cluster analysis

Each one of the eight time series was resampled to 1Hz and was processed using a Moving Window (MW), with a width ranging from 3 to 100 consecutive data. Six statistics were calculated for each window: the mean (M), standard deviation (S), interquartile range (IQR), kurtosis (K), obliquity or skewness (O) and coefficient of variation ($C = S/M$). In this way, 48 new statistical time series were obtained for each patient, given a window width.

From these time series, the optimal width of the MW was selected in the range from 3 to 100 by using two U Mann-Whitney tests between groups S and F and between groups S and R. Total p -values of the comparison between these groups, for all statistical time series was obtained. Minimal local and global values were obtained for a window width of 8 and 55 respectively (Fig. 1). We kept for further analysis the time series for the 48 features of each patient corresponding to these optimal (local minimum of 8 and global minimum of 55) window widths. Hence, a sequence of approximately 1,800 data points (or patterns) of dimension 48 were available for each patient and window widths of 8 and 55.

From these patterns, a data cluster analysis was carried out into each patient and window width, using the k -means method to determine the best number of clusters. Patient data were split into k mutually exclusive clusters (minimum 2, maximum 10). The partitions are assigned and defined by the data within each cluster and their centroids. The centroid of a cluster is considered to be the point at which the sum of the distances to it of the data points in the cluster is minimal, using the Fisher statistic.

The greater intra-cluster cohesion and inter-cluster separation is obtained, for a given k , by minimising the sum of distances between centroids and data points in their clusters. To avoid local minima results, the minimization

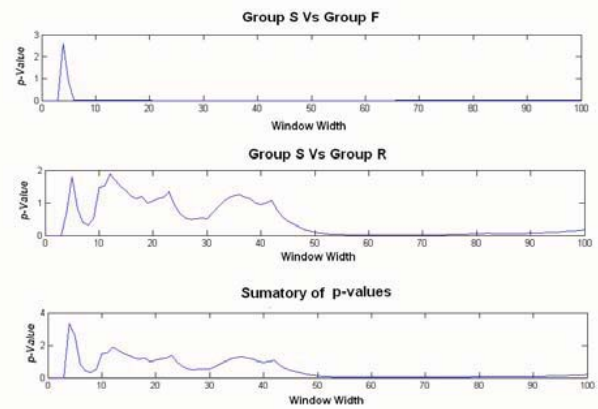


Fig. 1. Total p -values of the comparison between groups S-F, between groups S-R, and total sum of p -value, for each characteristic of each patient and all windows width.

process must be made several times to get the best clustering.

As a result of this process, it was obtained that most data points from each patient were grouped into two big conglomerates, with a dominant cluster containing the majority of points. Figure 2 shows as an example a patient of each study group, for a moving window of width 8. Each graph contains (a) a representation of the silhouette of each cluster, which represents the pooled data with the corresponding number of clusters, and (b) the average values of the cluster silhouette for different number of clusters ranging from 2 to 10.

The silhouette graph represents the data closeness in a cluster. Values close to +1 represent the greatest distances with the other clusters, some values close to 0 do not distinguish between a cluster and another, and negative values (-1) indicate data erroneously assigned to its cluster. The highest average value of the silhouette represents the number of clusters with greater intra-cluster cohesion and greater inter-cluster distance.

Finally, a statistical analysis was made to determine the variables that showed statistically significant differences between the study groups S, F and R, considering the two clusters that contained the most information, since the cluster analysis had shown that these two clusters covered the vast majority of data. Tables 1 and 2 present the average percentage represented by the main cluster (C1) and the two larger clusters (C1 + C2) for each group of patients, considering window widths of 8 and 55 respectively.

TABLE I
AVERAGE PERCENTAGE OF DATA REPRESENTED BY THE MAIN CLUSTER (C1) AND THE TWO LARGER CLUSTERS (C1 + C2) CONSIDERING A WINDOW WIDTH OF 8 FOR EACH GROUP OF PATIENTS

Window Width 8	C1	C1 + C2
Group S	69.16 ± 13.16	98.08 ± 7.69
Group F	68.86 ± 14.44	98.60 ± 4.70
Group R	69.83 ± 10.93	100

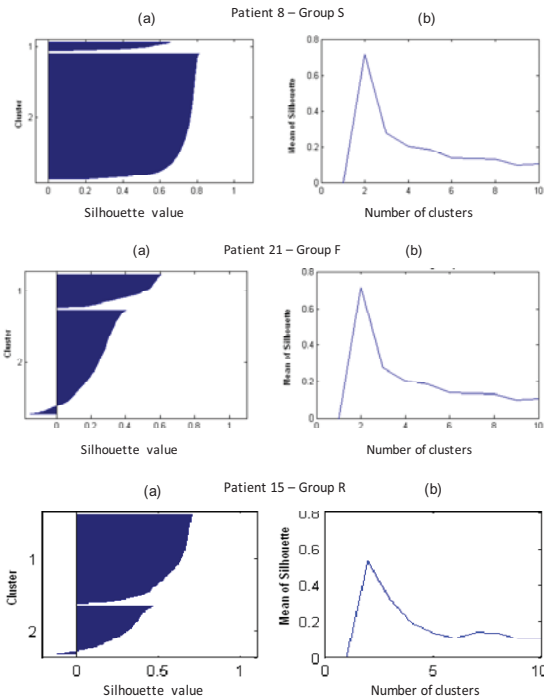


Fig. 2. (a) Silhouette of clusters, representing the appropriated pooled data according to the number of clusters for each patient, and (b) the number of clusters based on the average value of the silhouette, in accordance with the intra-cluster cohesion and the increased inter-cluster distance, corresponding to a patient in group S, a patient in group F and a patient in group R.

TABLE II
AVERAGE PERCENTAGE OF DATA REPRESENTED BY THE MAIN CLUSTER (C1) AND THE TWO LARGER CLUSTERS (C1 + C2) CONSIDERING A WINDOW WIDTH OF 55 FOR EACH GROUP OF PATIENTS

Window Width 55	C1	C1 + C2
Group S	58.48 ± 20.08	87.02 ± 21.42
Group F	61.05 ± 22.53	87.50 ± 21.34
Group R	60.56 ± 19.88	89.95 ± 19.58

According to the results of cluster analysis, the highest concentration of information was obtained with a window width of 8, while for a width of 55 a greater dispersion of data was found. Selecting the window width of 8, a single representative value of each variable for each cluster (C1 and C2) was calculated as the average value of the data point values in the cluster, thus obtaining two centroids for each patient, i.e. two vectors representing the two clusters.

Tables 3 and 4 show the variables with statistically significant differences between groups S – F and F – R, obtained by the U test of Mann-Whitney. The comparison between groups S – R did not find any variable with statistically significant differences.

B. Neural networks and feature selection

We have carried out three series of experiments for the discrimination of patients in two classes, weaning success (S) and failure (F), from the collected data of patients in the

TABLE III
VARIABLES SHOWING STATISTICALLY SIGNIFICANT DIFFERENCES BETWEEN GROUP S AND GROUP F ($P < 0.05$)

Group S vs. Group F	
Variable	p-values
M (T_{Tot})	<0.0005
M (T_I)	0.002
M (T_E)	<0.0005
S (T_E)	0.027
IQR (T_E)	0.018

TABLE IV
VARIABLES SHOWING STATISTICALLY SIGNIFICANT DIFFERENCES BETWEEN GROUP F AND GROUP R ($P < 0.05$)

Group F vs. Group R	
Variable	P-values
M (T_{Tot})	0.033
M (T_I)	0.023
M (T_E)	0.019
IQR (T_E)	0.015

S and F groups using feed-forward neural networks as classifiers (both linear single-layer perceptrons and non-linear two-layer perceptrons). As result of the data pre-processing and cluster analysis described previously, there were approximately 1,800 patterns of 48 variables and 2 centroids, corresponding to the main and secondary clusters, available for each patient. The classification baseline for comparison is 69.8% correctness, which is the percentage of patients in the majority class (group S).

In the first series of experiments, 10-fold cross-validation (CV) was performed on the whole set of patterns (226,800 patterns), training both linear and non-linear neural nets. The average pattern classification correctness in the test sets did not exceed the classification baseline in any of the different trials.

In the second series of experiments, we used only the cluster centroids instead of the whole set of data points. Again 10-fold CV was performed, training both linear and non-linear neural nets, in three cases, using for each patient (a) the two centroids, (b) only the centroid of the main cluster and (c) only the centroid of the secondary cluster. The CV results that were obtained taking into account the full set of 48 variables are displayed in Table 5.

Then, for each one of the six combinations (linear/non-linear net and a/b/c cases of data selection), feature selection was performed by carrying out several runs of a sequential backward selection method in order to determine a good subset of input variables. Features were eliminated one by one, selecting them in the way described in [11] for non-linear nets and in the way described in [12] for linear nets. The network was retrained each time a feature was removed. Once identified the feature subset providing the best result for each combination, a new 10-fold CV was performed to obtain the final average test classification

result using the selected variables. These final results are shown in Table 6.

TABLE V

CROSS-VALIDATION CLASSIFICATION RESULTS (CORRECTNESS AVERAGE PERCENTAGE) FOR SIX COMBINATIONS OF DATA SELECTION AND NEURAL NETS USING THE FULL SET OF 48 VARIABLES

Type of neural net Data selection	Linear perceptron	Non-linear 2-L perceptron
Two cluster centroids	64.4	67.0
Main cluster centroids	72.0	71.2
Secondary cluster centroids	63.2	66.2

TABLE VI

CROSS-VALIDATION CLASSIFICATION RESULTS (CORRECTNESS AVERAGE PERCENTAGE) FOR SIX COMBINATIONS OF DATA SELECTION AND NEURAL NETS USING THE BEST SUBSET OF VARIABLES OBTAINED FOR EACH CASE AFTER SEQUENTIAL BACKWARD SELECTION

Type of neural net Data selection	Linear perceptron	Non-linear 2-L perceptron
Two cluster centroids	76.4	77.1
Main cluster centroids	84.0	82.7
Secondary cluster centroids	78.4	78.7

In the third series of experiments, the neural networks that performed the best in the previous experiments (84.0% total correctness, 79.3% group S correctness and 86.1% group F correctness), i.e. linear perceptrons trained from main cluster centroids using only a selected subset of 19 variables ($K(T_I)$, $C(V_T/T_I)$, $M(V_T)$, $K(V_T/T_I)$, $S(RR)$, $M(V_T/T_I)$, $C(RR)$, $M(T_I/T_{Tot})$, $O(T_I)$, $O(V_T/T_I)$, $O(T_I/T_{Tot})$, $I(T_I)$, $K(T_{Tot})$, $S(V_T/T_I)$, $C(T_{Tot})$, $I(T_I/T_{Tot})$, $I(V_T)$, $S(T_{Tot})$, $M(f/V_T)$) were applied to classify the set of patterns. In order to discriminate the patients from the classification of their patterns, the following decision rule was used: a patient was classified as S (respectively F) if more than 50% of his/her patterns were classified as S (respectively F) by the neural net. The results came down to 74.6% total correctness, 97.7% group S correctness but only 21.1% group F correctness. As a last attempt, we changed the decision rule by increasing up to a 70% the percentage of S-labelled patterns needed to classify a patient in the class S, with the aim of improving the classification of class F. The new results were balanced (69.1% total correctness, 69.3% group S correctness and 68.4% group F correctness) but very close to the baseline and considerably worse than the results obtained using only the centroids of the main clusters as test.

IV. DISCUSSION AND CONCLUSIONS

The main objective of this work was to analyse the respiratory pattern variability in patients on weaning trials from mechanical ventilation, using cluster analysis and neural networks. With data mining process is possible to identify trends within data that go beyond simple data analysis.

According to the results of cluster analysis, the highest

concentration of information was obtained with a window width of 8, whereas window width of 55 presented a greater dispersion of data.

The best number of clusters for each patient and window width was obtained using the k -means method. The most data points were grouped into two big conglomerates. A dominant cluster contained around 69% of information (window width 8). With the two larger clusters we obtained 98% (average percentage) of information of groups S and F, and 100% of group R. The best statistical differences were obtained comparing groups S – F and F – R, and with the mean value of the breath duration.

Applying neural networks, the best result was obtained considering a linear single-layer perceptron and main cluster centroids, with 19 of 48 features (84% total well classified).

Many variables for successful outcomes from mechanical ventilation have been identified, but there are not specific and reproducible criteria clearly established. It should be validated with a larger number of patients.

ACKNOWLEDGEMENTS

The authors would like to thank to Dr. A. Ballester and his collaborators of the Hospital Universitario de Getafe, Spain, by their collaboration in the signal database acquisition.

REFERENCES

- [1] M. Meade, G. Guyatt, D. Cook, L. Griffith, "Predicting success in weaning from mechanical ventilation", *Chest*, 120, 6, 2001.
- [2] M.J. Tobin, "Advances in mechanical ventilation", *N. Engl. J. Med.*, Vol. 344, N. 26, pp. 1986-1996, 2001.
- [3] G. Benchetrit, "Breathing pattern in humans: diversity and individuality", *Respiration Physiology*, 122, pp. 123-129, 2000.
- [4] P. Caminal, L. Domingo, B.F. Giraldo, M. Vallverdú, S. Benito, G. Vázquez, D. Kaplan, "Variability analysis of the respiratory volume based on nonlinear prediction methods", *Medical & Biological Engineering & Computing*, 42, pp. 86-91, 2004.
- [5] M.J. Tobin, M.J. Mador, S.M. Guenter, R.F. Lodato, M.A. Sackner, "Variability of resting respiratory center drive and timing in healthy subjects", *J. Appl. Physiol.*, 65, pp. 309-317, 1988.
- [6] E.N. Bruce, "Measures of respiratory pattern variability", in *Bioengineering approaches to pulmonary physiology and medicine*, Plenum Press, pp. 149-160, 1996.
- [7] M.C. Khoo, "Determinants of ventilatory instability and variability", *Respiratory Physiology*, 122, pp. 167-182, 2000.
- [8] Howard E.A. Tinsley, Steven D. Brown, Edit. "Handbook of applied multivariate statistics and mathematical modelling". Ed. Academic Press. 2000.
- [9] C.M. Bishop, "Neural Networks for Pattern Recognition", Oxford University Press, 1995.
- [10] A. Gottschalk, M.C. Hyzer, R.T. Geer, "A comparison of human and machine-based predictions of successful weaning from mechanical ventilation", *Med. Decis. Making*, 20(2): 160-9, 2000.
- [11] E. Romero, J.M. Sopena, "Performing feature selection with multi-layer perceptrons", *IEEE Transactions on Neural Networks* 19 (3), 431-441, 2008.
- [12] P.J.G. Lisboa, E. Romero, A. Vellido, M. Julià-Sapé, and C. Arús, "Classification, dimensionality reduction, and maximally discriminatory visualization of a multicentre 1H-MRS database of brain tumours", *Proc. 7th International Conference on Machine Learning and Applications*, 613-618, 2008.